# Recent Trends in Deep Generative Models: a Review

**Conference Paper** · September 2018

DOI: 10.1109/UBMK.2018.8566353



2 authors:

Ceren Guzel Turhan
Gazi University


H. Ş. Bilge
Gazi University

# Recent trends in deep generative models: a review

Ceren GUZEL TURHAN
*Computer Engineering Department*
*Faculty of Engineering*
*Gazi University*
Ankara, Turkey
cerenguzel@gazi.edu.tr

Hasan Sakir BILGE
*Electric-Electronic Engineering Department*
*Faculty of Engineering*
*Gazi University*
Ankara, Turkey
bilge@gazi.edu.tr

***Abstract*—With the recent improvements in computation power and high scale datasets, many interesting studies have been presented based on discriminative models such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) architectures for various classification problems. These models have achieved current state-of-the-art results in almost all applications of computer vision but not efficient sampling out-of-data, understanding of data distribution. By pioneers of the deep learning community, generative adversarial training is defined as the most exciting topic of computer vision field nowadays. With the influence of these views and potential usages of generative models, many kinds of researches were conducted using generative models especially Generative Adversarial Network (GAN) and Autoencoder (AE) based models with an increasing trend. In this study, a comprehensive review of generative models with defining relations among them is presented for a better understanding of GANs and AEs by pointing the importance of generative models.**

***Index Terms*—Generative models, adversarial training, generative adversarial network, discriminative models, autoencoders**

## I. INTRODUCTION

Content generation is a topic with an increasing trend due to potential usages. With the recent improvements on adversarial learning concept on neural networks, generative models take many researchers' attention due to obtained promising results. Apart from CNN, RNN based discriminative models, which take consideration of conditional probability and work well for representation learning but not efficient to predict out-of-samples, out-of-samples can be generated to reject or sample with generative models. For predicting out-of-samples, generative models based on the joint probability of input pairs will be meaningful.

Nowadays, adversarial networks attract the attention of researchers through the discovery of adversarial examples and their effects on neural networks. Adversarial examples are obtained by manipulating original images via perturbations. These manipulations could not be seen easily on images but cause different predictions. Adversarial examples may not be commonly seen in practice but adversarial trained networks will be more robust and will be performing well at the same time [20].

Until now, unsupervised autoencoder and adversarial learning based generative models are designed for generating synthetic contents. In this study, generative models are reviewed to understand recent trends in the deep learning society. Generative models are categorized into 5 categories according to model architecture to specify relations within them. Moreover, a relational evaluation of generative models is presented in this paper. The idea behind this review is to summarize recent models based on associations among them according to their applications in literature and to highlight researchers about this increasing interest.

## II. THE IMPORTANCE OF GENERATIVE MODELS

Generative models are commonly studied for the last three years for their capability of generating data not only for estimating density. Generative models are important for

- Generating synthetic but realistic images. The following section focuses on generative models for this task. This review aims to collect generative models for mostly image synthesizing.
- Generating contents with predefined words and sentences as in [44], [65].
- Adversarial training that is significant to train models with adversarial samples to improve and assess model classification ability such as [45].
- Completing missing parts of data. Missed data is a problem for evaluating model correctly. With generative models, it is possible to predict missing parts of data because generative models give an idea about data manifold that is called as manifold learning. This is also called an image inpainting or completion with their models [62], [27], [60], [63] for image processing society.
- Manipulating original images based on predefined features. Generative models are capable of manipulating images based on latent codes. Not only pose, image parts, objects such as [16], [43], [7] but also image scenes can be altered using generative models that is defined as image-to-image translation. [37], [30], [67] models are able to map images from domain A to B.
- Working with multi-modal outputs. Generative models can be used for many tasks with a single input without separately training as in [23], [56].

- More samples from the same distribution. This is important for many tasks to obtain better models such as [1], [58] having fewer samples.
- Enhancing data quality. It is seen in review studies that generative models can be used to enhance data quality such as generating super-resolution images [36], [47] [59].

## III. Generative Models

In this paper, deep generative models are handled to clarify recent trend in the deep learning society. Among them, parametric models are associated amongst themselves to show a brief literature summary in Figure 1.

The recent studies of the generative model are categorized into unsupervised fundamental models, AE based models, autoregressive models, GAN based models and AE-GAN hybrid models to associate generative models easily.

### *A. Unsupervised fundamental models*

Unsupervised fundamental models were studied well for handwritten digit classification and texture synthesis tasks but have not been sufficient due to their blurry results. Restricted Boltzmann Machine (RBM) [49], which is a type of Boltzmann Machines (BM) has been applied various problems with improvements on computational powers. RBM and Deep Boltzmann Machine(DBM) [48] can reconstruct input image from its latent representations using their generative decoders with Gibbs sampling [9]. Markov chain Monte Carlo (MCMC) [18] is a method for generating fair samples using random dice in RBMs. With the stacked version of RBM, which is called as Deep Belief Networks (DBN) [25], it acts as a multilayer learning architecture providing features from high-level representations [14].

### *B. AE based models*

AE [64] combines two networks: encoder and decoder in a single network for unsupervised learning. The decoder part acts a generator since it generates a representation of data from compressed data. Stacked Autoencoder (SAE) [3] is suggested to train autoencoder with the greedy layerwise approach in a stacked manner. Denoising Autoencoder (DAE) model vincent2008extracting is introduced for robust representation learning giving corrupted data as input. Stacked Denoising Autoencoder (SDAE) [55] is an extension of DAE and SAE models. To form the SDAE model, DAE is stacked in a single network. Due to the overfitting and gradient descent problem, a variational component is proposed to model Variational Autoencoder (VAE) [32]. In the study [31], a deep generative model based on variational methods is presented in a semi-supervised manner, that we refer semi-supervised VAE as SS-VAE for the rest of the paper. to generate digits. Generative Stochastic Network (GSN) [52] is extended form of generalized DAE (GDAE) [4] to generate images by using learning parameters as in one-step of a generative MCMC. Adversarial Autoencoder (AAE) [39] is another autoencoder based generative model combining GAN objective that is adversarial training replacing KL-divergence term. In addition to the classical autoencoder network, the other network receives samples from prior distributions and latent code distributions to distinguish them. They projected MNIST classes to the latent space then generated realistic images on this manifold.

Deep Recurrent Attentive Writer (DRAW) [21] is an attractive work about recurrent autoencoder using attention mechanism. Because of its recurrence structure, multiple autoencoders were comprised. Owing to attention mechanism, the model can decide to be focused on which parts of images and to be drawn on which parts of the output.

Denoising Variational Autoencoder (DVAE) [28] combines VAE and DAE approaches to get robust model injecting noises not only input level but also in a stochastic hidden layer as in VAE. Due to the limited disentangled performance of VAE, $\beta$ Variational Autoencoder ($\beta$-VAE) [24] is proposed with additional $\beta$ hyperparameter to learn disentangled representations by balancing latent code capacity and independence generative factors.

Wasserstein Autoencoder (WAE) [53] is another AE based model for minimizing loss based on Wasserstein distance using proposed regularizer.

### *C. Autoregressive models*

In this part, autoregressive models which model images pixel-by-pixel rather than whole image are handled.

MADE [17] is a modified autoencoder network using the autoregressive property to estimate any distribution from a set of examples. In this model, there are skipped connections and multiplicative binary masks to ensure autoregressive property.

PixelCNN Decoder [54] is an autoregressive generative model based on CNNs. It acts as a decoder of an autoencoder. PixelCNN models the conditional distribution of seen pixel values. A gated CNN is used to remember prior pixel values in this gated architecture. In this model, a latent factor term is used to generate images. An RNN version, PixelRNN, is also proposed in the same study. In the following study of RixelRNN model by [42], RixelRNN is used including 12 LSTM layers adopting the convolutional approach. They also defined Row LSTM, Diagonal BiLSTM using residual connections among LSTM layers and multi-scale version of PixelRNN. A modified version of PixelCNN is proposed as PixelCNN++ [51] to simplify the architecture while improving the performance. In the modified version, they used discretized logistic mixture likelihood on whole pixels using some tricks as downsampling, dropout, and skip-out connections obtaining better results based on the log likelihood.

In the study [22], a latent variable model called as PixelVAE is proposed to combine the benefits of VAEs and PixelCNNs in a model. A conditional PixelCNN is used for the output of VAE's decoder in this model.

Variational Lossy Autoencoder(VLAE) [8] is proposed combining VAE with autoregressive models to improve modeling performance of VAE.

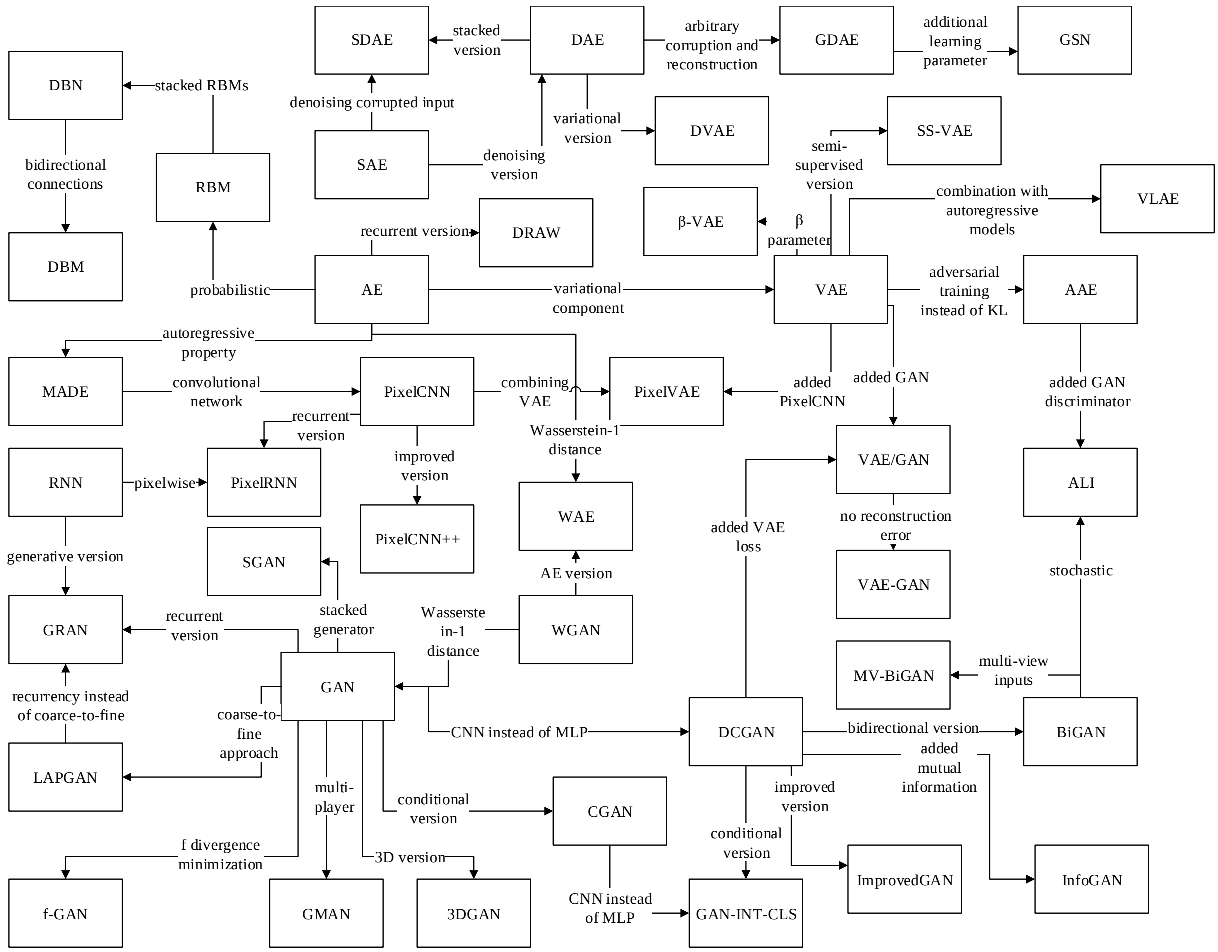


Fig. 1. The relation diagram of generative models

### *D. GAN based models*

Before GAN [19] models, generative models have already existed, as mentioned before, but they could not attract much attention due to not only computational issues but also their complicated structures.

GAN brings a new breath to the deep learning community. And now, deep learning with adversarial training seems as one of the most robust technique. With generative adversarial networks, not only a good classifier based on neural network, which will be named as a discriminative network in the rest of this paper, but also a generative network for generating realistic adversarial samples are designed a single architecture. That is, now, we have a network that is aware of internal representations due to training for distinguishing real and artificial inputs. Furthermore, we have a tool that can generate artificial but realistic samples. Having such a tool is interpreted as one of the todays most interesting topics due to potential usages of content generation.

In GAN, the model consists of two networks: the generator ($G$) and the discriminator ($D$) network. In there, $G$ and $D$ networks are multi-Layer perceptrons (MLPs). In the model, the GAN training procedure is completed separately, but simultaneously. The GAN model takes a noise input $z$ which is defined as prior probability $p_z$ then tries to learn the distribution of generator, $p_g$, by representing a mapping $G(z;\theta_g)$ from $z$ to data space. The discriminator network $D$ takes an input image, x then finds a mapping $D(x;\theta_d)$ from $x$ to a single scalar, that is the probability of the image $x$ from $p_{data}$. $p$ data defines where images are sampled from. The output of network $D$ returns a value close to 1 if the $x$ is a real image that is from $p_{data}$. Otherwise, if the $x$ is from $p_g$, the output will be very close to 0. The main goal of the network $D$ is maximizing $D(x)$ for an image from true data distribution, $p_{data}$ while minimizing $D(x) = D(G(z;\theta_g)$ for generated images from $p_z$ not p data. The aim of the generator $G$ is to fool the network $D$, that is maximizing $D(G(z;\theta_g))$. This is equivalent to minimize $1 - D(G(z;\theta_g))$ because the $D$ is a binary classifier. There is a conflict among these aims which is called as minimax game given in equation 1.

$$\min\max E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1 - D(G(z;\theta_g)))] \quad (1)$$

The global optimum of this minimax game is the case of $p_g = p_{data}$.

After GAN model, Conditional Generative Adversarial Network (CGAN) [16] has been proposed as an extension of the original GAN model to generate facial images based on conditioning features i.e. facial attributes. With this extension, generator model can generate an image with a specified feature and similarly discriminative model can understand whether input image contains an object with specified features or not.

Laplacian Pyramid of Generative Adversarial Network (LAPGAN) [9] as an extension of GAN has been proposed including multiple pairs of $G$ and $D$ networks. In the beginning, the network $G$ samples very small sized images taking a noise vector as input. Then, it learns to generate upscaled images. This scaling from coarse to fine acts as a Laplacian pyramid. In the discriminative part, $D$ takes an upscaled image and either $G$s difference image or difference images from the training set as input. The network $D$ consists of convolutional layers with a sigmoid unit differing from the original GAN approach.

Deep Convolutional Generative Adversarial Network (DCGAN) [43] is a convolutional GAN based architecture for generating images. In DCGAN, both network $G$ and $D$ are convolutional. In network $D$, strided convolutional layers were preferred instead of pooling layer. Leaky ReLUs were performed as activation functions. Their other contributions are the visualization of representations, activating/deactivating some of the feature maps and word-to-vector based arithmetic on face images.

GRAN [29] is a GAN model based on a sequential process similar to DRAW and LAPGAN models. The network $G$ generates images in multiple time-steps beginning with an empty plane $\Delta C_1$ and noise vector $z_0$. Then, in the following time-step, it takes the output of previous time-step and noise vector $z_1$. Until the end of all time-steps, it concatenates the output of previous time-steps to generate an image. In addition to this new architecture, a battle between GANs approach was suggested to evaluate the quality of generated images.

Improved GAN [50] is an impressive study among GAN-based models. This study includes some manipulations on GAN training scheme. These are the usage of different objective approach for back-propagation, mini-batch, punishing weights, the usage of smoothed labels instead of binary labels, virtual batch normalization, and a novel inception scoring.

In the work [10], Bidirectional GAN (BiGAN) is proposed to map data $x$ to latent code $z$ as in autoencoder and similar to the [12] learning with inverse mapping. To simply learning in multi-view data, Multi-view BiGAN (MV-BiGAN) is proposed by [6] to estimate density and then predicting additional views.

To extend two-player minimax in GANs to multi-player, a Generative Multi-Adversarial Network (GMAN) is proposed in the study [13]. To accelerate training with multiple discriminators, they updated objective function according to multi-player minimax game.

Stacked Generative Adversarial Network (SGAN) [26] consists of a bottom-up pretrained encoder and a top-down stack of the generator for generating low-level representations using high-level representations.

In a divergence perspective, f-divergence Generative Adversarial Network (f-GAN) [41] is proposed for f-divergence minimization based on density ratio estimation in generative part. In this study, GAN is defined as a special case of f-GAN. Similarly, Wasserstein Generative Adversarial Network (WGAN) [2] is introduced as one of the popular GAN models due to its diminishing effect on mode collapse problem meanwhile reducing instability in training. They suggested an Earth-mover distance (Wasserstein-1 distance) as loss and performance metric for extracting and sampling from true probability distributions.

Information Maximizing Generative Adversarial Net (InfoGAN) [7], a variation of the GAN model, depends on mutual information in the generative part. In this model, $G$ generates an image by a function of $G(z, c)$ where $c$ is a latent code and could be more than one variable with different distributions. They expressed a relation between image, $x$, and $c$ code. Using this relation, they updated loss function of GAN model based on mutual information.

Another attractive study going one further step in terms of image generation is GAN-INT-CLS [44]. In this model, images are generated in terms of a given sentence, that is, it converts text description to images. GAN trains conditional text features obtained by a recurrent encoder. Using concatenation of encoded text and noise vector, images are generated to pass from convolutional layers until the getting $4 \times 4$ dimensional data. Then, the network realizes a depth concatenation with encoded text.

In addition to previous studies for 2D image generation, generative models are extended to generate 3D objects. These models, generally generate 3D objects using 3D input objects in a supervised manner (CAD models etc.) while some of them are trained to learn predictions instead of objects without using any 3D data during training. 3D object generator in an unsupervised manner and reconstructing 3D objects from 2D images are handled in three-dimensional Generative Adversarial Networks (3DGAN) [57]. In addition to unsupervised shape and viewpoint generation, Projective Generative Adversarial Network (PrGAN) [15], as a composed of 3DGAN, viewpoint generator, projection module and 2D discriminator, proposed to generate shape and viewpoint in an unsupervised manner.

### *E. AE-GAN hybrid models*

A group of study is conducted to hide GAN's lack of inference mechanism with an encoder network. One of the interesting studies about VAE and GAN [34] combine these two approaches in one model which is called as VAE/GAN. Learned discriminative features in GAN have been used as VAEs reconstruction objective instead of element-wise reconstruction objective. This model can enable to learn encoding, generating and finally discriminating. The loss function of VAE/GAN is updated by the summation of VAEs prior loss, feature similarity based log likelihood and the loss of GAN model. Moreover, a simple arithmetic using high-level features also realized to generated images with specified features. Another VAE-GAN hybrid study by [33] is about pretrained auxiliary network loss usage instead of reconstruction error in VAE. Similarly, [11] used VAE with GAN based on their DeePSiM loss function to prevent blurry reconstructed images. In [12]'s ALI model, a GAN is employed with an adversarial autoencoder model. In this approach, there is no any explicit reconstruction error for the optimization of AAE. Moreover, GAN's discriminator network takes input pairs $(x, z)$ instead of only latent code $z$.

3DGAN model is combined with VAE on the same study as 3D-VAE-GAN [57] to generate 3D objects using encoded 3D input data.

## IV. Comparison of generative models

Since now, generative models are defined for image generation task. Actually, generative models are suitable for various tasks such as super-resolution, image colorization, image inpainting, image manipulation, image generation, text to image generation, image to image generation, etc. Before generation task, autoencoders were realized image reconstruction and then they are taken as appropriate for the generative task due to their decoder network. Enhancing image resolution is another task for generative models such as SRGAN [35]. Inpainting or image completion task is achieved by generative models as [61]. Manipulating images is also possible using generative models as in InfoGAN [7], CGAN [16], and [66]. Generating images from words as in [44] and applying some arithmetic [43], that is from image to image generation, are impressive applications of generative models. Other generative model applications are one-shot generalization [46], specified input generation [40], image colorization [5], domain adaptation such as CoGAN [38], etc.

To analyze the quality of generated images are not easy and it is already an open question in generative models. As an evaluation metric, the log-likelihood metric is generally preferred to evaluate generated contents via generative models. In some studies such as GAN, a single neighborhood, inception score and classification accuracy are chosen to evaluate generated images. In LAPGAN, human inspections evaluate generated images. The classification performance is another evaluation metric for generative models as in [66] using extracting features from models. In [29], the bottle between GANs approach is proposed to evaluate generative models.

## V. Conclusion and Discussion

Nowadays, deep learning based models achieve current state-of-the-art results in almost all applications. It is seen that the generative models, especially including adversarial training, are determined as the most interesting topic for the deep learning community. It is clear to see that the majority of recent studies, especially from the last three years, is about generative models with an increasing trend. Due to this increasing interest, there is a demand to gather recent trends about generative models. In this comprehensive study, recent generative models are given to collect all approaches and answer some of the following questions: what does deep learning society study about and why? what are the application areas and other trends in generative models? A relational evaluation of generative papers is also present to give any idea about the relevance of studies. As a result, it can be inferred from this review that there is an increasing interest in generative models due to their potential usages and importance for computer vision community.

## References


[1] A. Antoniou, A. Storkey, and H. Edwards. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017.

[2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[3] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, pages 153–160, 2007.

[4] Y. Bengio, L. Yao, G. Alain, and P. Vincent. Generalized denoising auto-encoders as generative models. In *Advances in Neural Information Processing Systems*, pages 899–907, 2013.

[5] Y. Cao, Z. Zhou, W. Zhang, and Y. Yu. Unsupervised diverse colorization via generative adversarial networks. *arXiv preprint arXiv:1702.06674*, 2017.

[6] M. Chen and L. Denoyer. Multi-view generative adversarial networks. *arXiv preprint arXiv:1611.02019*, 2016.

[7] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2172–2180, 2016.

[8] X. Chen, D. P. Kingma, T. Salimans, Y. Duan, P. Dhariwal, J. Schulman, I. Sutskever, and P. Abbeel. Variational lossy autoencoder. *arXiv preprint arXiv:1611.02731*, 2016.

[9] E. L. Denton, S. Chintala, R. Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015.

[10] J. Donahue, P. Krähenbühl, and T. Darrell. Adversarial feature learning. *arXiv preprint arXiv:1605.09782*, 2016.

[11] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. *arXiv preprint arXiv:1602.02644*, 2016.

[12] V. Dumoulin, I. Belghazi, B. Poole, A. Lamb, M. Arjovsky, O. Mastropietro, and A. Courville. Adversarially learned inference. *arXiv preprint arXiv:1606.00704*, 2016.

[13] I. Durugkar, I. Gemp, and S. Mahadevan. Generative multi-adversarial networks. *arXiv preprint arXiv:1611.01673*, 2016.

[14] A. Fischer and C. Igel. An introduction to restricted boltzmann machines. In *Iberoamerican Congress on Pattern Recognition*, pages 14–36. Springer, 2012.

[15] M. Gadelha, S. Maji, and R. Wang. 3d shape induction from 2d views of multiple objects. *arXiv preprint arXiv:1612.05872*, 2016.

[16] J. Gauthier. Conditional generative adversarial nets for convolutional face generation. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester*, 2014:5, 2014.

[17] M. Germain, K. Gregor, I. Murray, and H. Larochelle. Made: Masked autoencoder for distribution estimation. In *ICML*, pages 881–889, 2015.

[18] C. J. Geyer. Markov chain monte carlo maximum likelihood. 1991.

[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[20] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[21] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra. Draw: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*, 2015.

[22] I. Gulrajani, K. Kumar, F. Ahmed, A. A. Taiga, F. Visin, D. Vazquez, and A. Courville. Pixelvae: A latent variable model for natural images. *arXiv preprint arXiv:1611.05013*, 2016.

[23] K. Hausman, Y. Chebotar, S. Schaal, G. Sukhatme, and J. J. Lim. Multi-modal imitation learning from unstructured demonstrations using generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 1235–1245, 2017.

[24] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. 2016.

[25] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
[26] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie. Stacked generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, 2017.
[27] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (TOG)*, 36(4):107, 2017.
[28] D. J. Im, S. Ahn, R. Memisevic, Y. Bengio, et al. Denoising criterion for variational auto-encoding framework. In *AAAI*, pages 2059–2065, 2017.
[29] D. J. Im, C. D. Kim, H. Jiang, and R. Memisevic. Generating images with recurrent adversarial networks. *arXiv preprint arXiv:1602.05110*, 2016.
[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks.
[31] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems*, pages 3581–3589, 2014.
[32] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
[33] A. Lamb, V. Dumoulin, and A. Courville. Discriminative regularization for generative models. *arXiv preprint arXiv:1602.03220*, 2016.
[34] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*, 2015.
[35] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.
[36] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint*, 2017.
[37] M.-Y. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, pages 700–708, 2017.
[38] M.-Y. Liu and O. Tuzel. Coupled generative adversarial networks. In *Advances in Neural Information Processing Systems*, pages 469–477, 2016.
[39] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.
[40] A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, and J. Clune. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. In *Advances in Neural Information Processing Systems*, pages 3387–3395, 2016.
[41] S. Nowozin, B. Cseke, and R. Tomioka. f-gan: Training generative neural samplers using variational divergence minimization. In *Advances in Neural Information Processing Systems*, pages 271–279, 2016.
[42] A. v. d. Oord, N. Kalchbrenner, and K. Kavukcuoglu. Pixel recurrent neural networks. *arXiv preprint arXiv:1601.06759*, 2016.
[43] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
[44] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 3, 2016.
[45] H. Ren, D. Chen, and Y. Wang. Ran4iqa: Restorative adversarial nets for no-reference image quality assessment. *arXiv preprint arXiv:1712.05444*, 2017.
[46] D. J. Rezende, S. Mohamed, I. Danihelka, K. Gregor, and D. Wierstra. One-shot generalization in deep generative models. *arXiv preprint arXiv:1603.05106*, 2016.
[47] M. S. Sajjadi, B. Schölkopf, and M. Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 4501–4510. IEEE, 2017.
[48] R. Salakhutdinov and G. E. Hinton. Deep boltzmann machines. In *AISTATS*, volume 1, page 3, 2009.
[49] R. Salakhutdinov, A. Mnih, and G. Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798. ACM, 2007.
[50] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2226–2234, 2016.
[51] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *arXiv preprint arXiv:1701.05517*, 2017.
[52] E. Thibodeau-Laufer, G. Alain, and J. Yosinski. Deep generative stochastic networks trainable by backprop. 2014.
[53] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*, 2017.
[54] A. van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al. Conditional image generation with pixelcnn decoders. In *Advances in Neural Information Processing Systems*, pages 4790–4798, 2016.
[55] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(Dec):3371–3408, 2010.
[56] V. Vukotić, C. Raymond, and G. Gravier. Generative adversarial networks for multimodal representation learning in video hyperlinking. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pages 416–419. ACM, 2017.
[57] J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in Neural Information Processing Systems*, pages 82–90, 2016.
[58] Q. Xu, Z. Qin, and T. Wan. Generative cooperative net for image generation and data augmentation. *arXiv preprint arXiv:1705.02887*, 2017.
[59] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M.-H. Yang. Learning to super-resolve blurry face and text images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 251–260, 2017.
[60] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li. High-resolution image inpainting using multi-scale neural patch synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 3, 2017.
[61] R. Yeh, C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with perceptual and contextual losses. *arXiv preprint arXiv:1607.07539*, 2016.
[62] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5485–5493, 2017.
[63] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. *arXiv preprint*, 2018.
[64] R. S. Zemel. Autoencoders, minimum description length and helmholtz free energy. NIPS, 1994.
[65] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. *arXiv preprint*, 2017.
[66] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros. Generative visual manipulation on the natural image manifold. In *European Conference on Computer Vision*, pages 597–613. Springer, 2016.
[67] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*, pages 465–476, 2017.